AC

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 July 2002 (25.07.2002)

PCT

(10) International Publication Number
WO 02/057497 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US02/01651

(22) International Filing Date: 18 January 2002 (18.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/262,724 18 January 2001 (18.01.2001) US

(71) Applicant and
(72) Inventor: **MORRIS, David, W.** [US/US]; 2841 Emerald Bay Drive, Davis, CA 95616 (US).

(74) Agents: **SILVA, Robin, M.** et al.; Flehr, Hohbach, Test, Albritton & Herbert, LLP, 4 Embarcadero Center, Suite 3400, San Francisco, CA 94111-4187 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/057497 A2

(54) Title: HIGH THROUGH-PUT CLONING OF PROTOONCOGENES

(57) Abstract: The present invention provides a process of identifying protooncogenes using high-throughput provirus tagging (HPT), e.g., by recovering host/virus junction sequences from chimeric transcripts containing both host and virus sequences.

HIGH THROUGH-PUT CLONING OF PROTOONCOGENES

General Background

Cancer is the phenotypic manifestation of a complex biological progression during which cells accumulate multiple somatic mutations, eventually acquiring sufficient growth autonomy to metastasize. Although inherited cancer susceptibility alleles and epigenetic factors influence the process, carcinogenesis is fundamentally driven by somatic cell evolution (*i.e.*, mutation and natural selection of variants with progressive loss of growth control). The genes which are the targets of these somatic mutations are classified as either protooncogenes or tumor suppressor genes, depending on whether their mutant phenotyes are dominant or recessive, respectively.

In several animal models, an important source of protooncogene somatic mutations is retrovirus infection. Retroviruses cause can cause cancer by essentially three mechanisms: (i) transduction of host protooncogenes (which then become viral oncogenes), (ii) *trans*-acting effects of viral gene products, or (iii) *cis*-acting effects of provirus integration on protooncogenes at or very near the site of integration. In the later case, only rare infected cells are affected. This phenomenon is called provirus insertion mutation, and will be discussed in detail in the following narrative.

As a normal consequence of the retroviral life-cycle, DNA copies of the retrovirus genome (called a proviruses) are integrated into the host genome. Accordingly, retroviruses are obligate mutagens. A newly-integrated provirus can affect gene expression in *cis* at or near the integration site by one of two mechanisms. Type I insertion mutations up-regulate transcription of proximal genes as a consequence of regulatory sequences (enhancers and/or promoters) within the proviral long terminal repeats (LTRs). These insertion mutations typically affect genes that are not expressed in the target tissue. Type II insertion mutations cause truncation of coding regions due to either integration directly within an open reading frame or integration within an intron upstream of the stop codon.

Provirus integration is random. Therefor, all host genes are targets of insertion mutation. In a chronically-infected tissue, a sufficient number of cells have new provirus insertions that, statistically, all genes in the genome are mutated. In rare cases, an insertion mutation will "activate" a host protooncogene, providing the affected cell with a dominant selective growth advantage *in vivo*. If the cell progresses to cancer, then the protooncogene insertion mutation will be present at clonal

stoichiometry in the tumor. Such "clonally-integrated" proviruses serve to "tag" the locations of protooncogenes in the genome. In cases where the proviral enhancer is responsible for dysregulation of the mutated protooncogene, the provirus can be 100 kb or more from the site of integration (but usually much closer).

This relatively tight linkage between clonally-integrated proviruses and protooncogenes is the basis for a classical experimental strategy, called "provirus tagging," in which slow-transforming retroviruses that act by an insertion mutation mechanism are used to isolate protooncogenes. The complete logic is as follows: (i) uninfected animals have low cancer rates, (ii) infected animals have high cancer rates, (iii) the retroviruses involved do not carry transduced host protooncogenes or pathogenic *trans*-acting viral genes, (iv) the cancer incidence must therefor be a direct consequence of provirus integration effects on host protooncogenes, (v) since provirus integration is random, rare integrants will "activate" host protooncogenes that provide a selective growth advantage, and (vi) these rare events result in new proviruses at clonal stoichiometries in tumors.

In contrast to mutations caused by chemicals, radiation, or spontaneous errors, protooncogene insertion mutations can be easily located by virtue of the fact that a convenient-sized genetic marker of known sequence is present at the site of mutation (*i.e.*, the provirus). Host sequences that flank clonally-integrated proviruses can be recovered using a variety of molecular techniques. Once these sequences are in hand, the tagged protooncogenes can be subsequently identified.

There are two unequivocal biological criteria that provide *prima facie* evidence that a protooncogene is present at or very near a proviral integration site. The first criterion is the presence of provirus at the same locus in two or more independent tumors. This is because the genome is too large for random integrations to result in observable clustering. Any clustering that is detected is indirect evidence for biological selection (*i.e.*, the tumor phenotype resulting from activation of a host protooncogene). The second criterion is a tumor with only a single insertion mutation. In this case, if there is only one insertion mutation, then that provirus is located at a protooncogene locus. If either of these criteria are met, sufficient evidence exists to reach a conclusion that a protooncogene locus has been located.

The provirus tagging concept has withstood two decades of testing in many retrovirus tumor models that have a provirus insertion mutation etiology. The biological logic is so compelling, and the experimental results so unequivocal, that the claim can be made that the activated genes are functionally-validated as protooncogenes at the time-of discovery. Formal confirmation typically involves

isolation of a full-length cDNA for use in a bioassay (either a cell-based transformation assay or transgenic mice).

Provirus tagging in avian and mammalian systems has led to the identification of approximately 50-60 protooncogenes (many of which were new genes not previously identified by other techniques). The three mammalian retroviruses that cause cancer by an insertion mutation mechanism are FeLV (leukemia/lymphoma in cats), MLV (leukemia/lymphoma in mice and rats), and MMTV (mammary cancer in mice).

Despite the tremendous promise of the provirus tagging approach, as originally designed it was not well-suited for large scale application. The main problem was that it was too laborious and, therefor, the risks of reisolating known genes became unacceptable for most investigators. As a consequence, the protooncogene discovery potential of this approach has remained largely untapped.

Recognizing this untapped potential, we designed and implemented HPT to overcome the limitations of the original provirus tagging approach (which were all fundamentally related to throughput). We were able to successfully increase provirus tagging throughput to the point where reisolation of known loci is no longer a problem. In fact, this is now a desirable outcome because it serves as an "internal control" that helps validate the biological relevance of the new genes that are recovered in parallel.

As a functional oncogenomics strategy, HPT has many advantages. First, it is a functional cloning rather than brute-force (*e.g.*, differential display-based)approach; and the genes that are recovered are functionally-validated at the time of discovery. Second, it has high biological relevance since protooncogenes are isolated directly from clinical material (rather than from cell lines, transplants, or materials generated by gene transfer). Third, it is amenable to automation, meaning that throughput and time-to-discovery is a simple function of research resources.

The invention is a process called high-throughput provirus tagging (HPT). HPT yields partial protooncogene cDNAs from retrovirus-induced tumors. Using these partial cDNAs, conventional techniques can be used to recover full-length cDNAs (we have not yet performed this final step). A conceptual diagram is shown in Appendix A and a flow chart of the process is shown in Appendix B.

HPT is derived from classical procedures for provirus tagging (see Appendix C for background information). It is specific for tumors induced by retroviruses that cause cancer via a provirus insertion mutation mechanism. This subset of retroviruses includes the mouse mammary tumor virus (MMTV). MMTV-induced tumors were used to implement the HPT process.

In tumors induced by provirus insertion mutation, new proviral integrants present at clonal stoichiometries tag the locations of host protooncogenes. The majority of such integrants fall outside of transcribed regions. However, a subset fall within sequences that are transcribed, and result in the formation of chimeric transcripts containing both host and virus sequences. HPT is designed to recover host/virus junction sequences from these chimeric transcripts.

The strategy used is a modified/optimized anchored-PCR (A-PCR)approach incorporating a custom anchor. The procedure amplifies host sequences upstream of 5' LTRs. If a transcript containing a host/virus junction is present in a tumor, then a unique fragment is generated by the A-PCR procedure, which can be detected by gel electrophoresis. In addition, one or more common fragments will be generated from retroviral transcripts that contain the 5' end of the 3' LTR.

The innovation that makes this approach feasible is that cDNAs are digested with a restriction enzyme with a 4 bp recognition sequence prior to amplification. This generates populations of target cDNAs that (1) have precise 5' ends, and (2) are sufficiently small to ensure that they will efficiently amplify. In addition, restriction enzymes are selected that produce the largest possible retroviral transcription products (so that they run at the top of the gel). This is critical because chimeric transcripts are present at much lower levels than the major retroviral transcripts. By selection of appropriate restriction enzymes, a large detection window is available in a region of the gel where the signal-to-noise ratio is most favorable. In addition, during amplification, cycling times are ramped to favor smaller products.

The provirus tagging strategy has been used for almost 20 years. It is a DNA-based detection method where identification of new genes requires positional cloning procedures to find genes adjacent to integration sites recovered from tumor DNA. This laborious process has been recently improved by PCR procedures. Nevertheless, unless the integration falls within known sequence, it is not possible to identify the affected gene without a large amount of additional work.

The advantage of HPT is that it is the first PCR-based provirus tagging approach that recovers protooncogenes from RNA. Because RNA is used, new protooncogenes are identified directly. Although only a fraction of tumors have insertion mutations that generate a chimeric transcript, the process has been designed to be high-throughput. As a consequence, the fact that most samples are non-informative is not a problem. In addition, the process is so efficient that recovery of know protooncogenes does not represent an unacceptable loss of effort, and, in fact, serves as an internal control to verify the robustness of the strategy.

I claim:

1. A method of identifying protooncogenes comprising:

inserting a provirus into the genome of a host forming a junction site of DNA of said virus and said host;

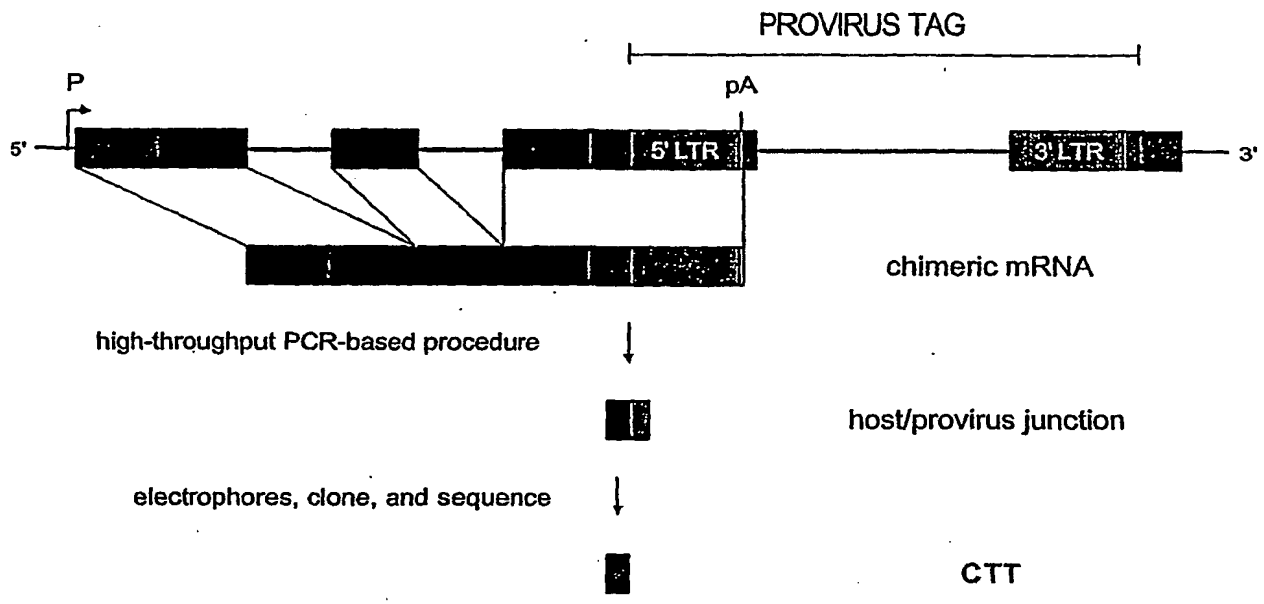
isolating mRNA from said host;

preparing cDNA from said mRNA;

amplifying said cDNA to identify the nucleic acid sequence of said junction site, whereby said candidate target gene is identified.

APPENDIX A

FIGURE 1 - CONCEPTUAL DIAGRAM



The HPT process can detect insertion mutations in either orientation anywhere in a transcribed sequence (including introns). Figure 1 illustrates an integration in the 3' untranslated region of a hypothetical protooncogene. This is the most common type of insertion mutation detected by HPT in the MMTV system.

APPENDIX B

FIGURE 2 - FLOW CHART

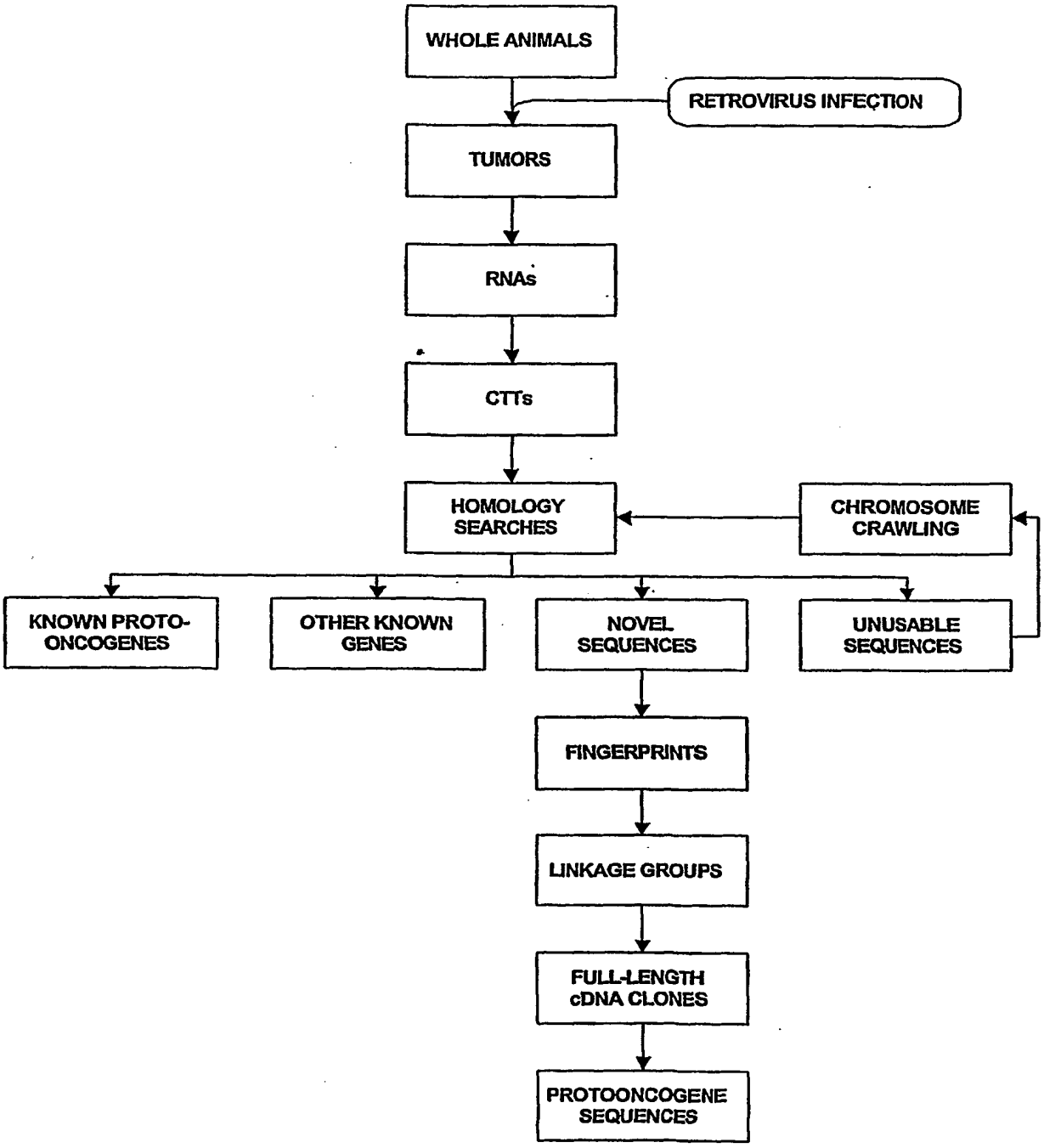
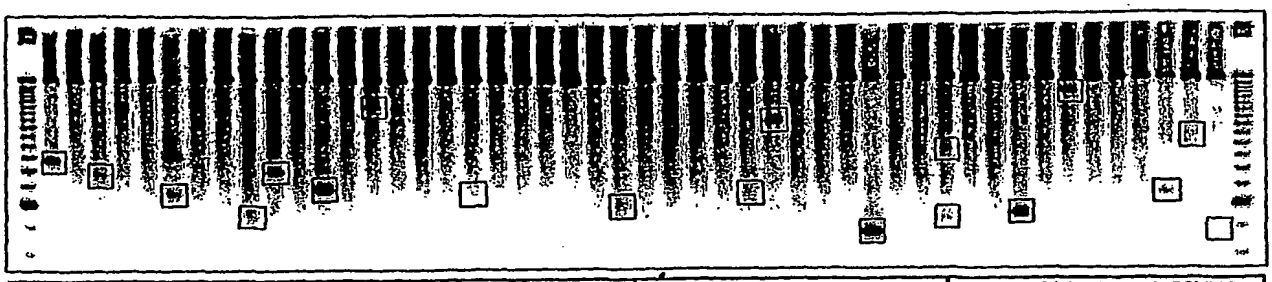


FIGURE 3 - EXAMPLE OF HPT SCREENING DATA



LANE	TUMOR	RESULT	LANE	TUMOR	RESULT	LANE	TUMOR	RESULT	LANE	TUMOR	RESULT
1	MM0001	*Wnt3*	13	MM0033	–	25	MM0046	–	37	MM0084	*Sp100*
2	MM0002	–	14	MM0034	*Fgf3*	26	MM0047	–	38	MM0085	–
3	MM0003	*Fgf3*	15	MM0035	–	27	MM0048	–	39	MM0091	–
4	MM0004	–	16	MM0036	–	28	MM0049	–	40	MM0092	*Wnt3a*
5	MM0005	–	17	MM0037	–	29	MM0064	novel	41	MM0093	–
6	MM0006	*Fgf8*	18	MM0038	*Myb*	30	MM0065	novel	42	MM0094	*Fgf3*
7	MM0007	–	19	MM0039	–	31	MM0066	–	43	MM0095	–
8	MM0008	–	20	MM0040	–	32	MM0074	–	44	MM0097	–
9	MM0009	*Ins2 /Igf2*	21	MM0041	–	33	MM0077	–	45	MM0099	–
10	MM0010	novel	22	MM0042	–	34	MM0078	unusable	46	MM0103	novel
11	MM0011	–	23	MM0044	–	35	MM0080	–	47	MM0134	*Wnt1*
12	MM0029	*Fgf3*	24	MM0045	unusable	36	MM0081	–	48	MM0250	unusable

APPENDIX D

A. General purpose or utility

HPT is new technology for isolating partial cDNAs representing functionally validated protooncogenes. It is a scaleable batch process that is amenable to high-throughput applications. Saturation mutagenesis of all accessible protooncogenes in the mammalian genome is feasible using HPT.

B. Brief description of the state of the art prior to your invention

The state-of-the-art prior to this invention was PCR-based isolation of provirus integration sites from tumor DNA. This method, as currently practiced, involves an inverse-PCR (I-PCR strategy). Identification of the activated protooncogene at a particular integration locus relies on prior characterization of the gene by other methods. For novel genes, positional cloning is required. The DNA-based method involves considerable risk, since it is not know until the end of the positional cloning process whether the locus under investigation is novel.

C. Technical description

1. Isolate total RNA from frozen tumor tissue.
2. Treat with DNase.
3. Prepare double stranded cDNA.
4. Digest with restriction enzyme.
5. Ligate anchor to digested cDNA.
6. PCR amplify targets with LTR and anchor primers.
7. Reamplify targets with nested LTR and anchor primers.
8. Electrophores amplification products.
9. Sample new band, if present.
10. Reamplify band.
11. Clone.
12. Determine sequence.
13. Assign CTT number.
14. Perform homology search.
15. If sequence is anonymous, design primers for fingerprinting.
16. Use primers to amplify BAC and YAC superpools.
17. Electrophores to determine banding pattern for the locus (fingerprint).
18. Assemble into linkage groups.

CTTs can be assembled into linkage groups based on their fingerprints. Using a representative CTT from each linkage group, conventional techniques can then be used to isolate full-length cDNAs for sequence analysis and deduction of the amino acid sequencing of the protooncogene.

D. Practicing the invention

Shown are the A-PCR results from an HPT analysis of 48 independent MMTV-induced tumors performed as described in the previous sections. Tabulated below the gel are the results of a BLAST search using the CTT from each recovered junction fragment (boxes). In addition to novel sequences, known targets of MMTV-insertion mutation were recovered. Also recovered, were known protooncogenes not previously recognized as targets of MMTV and know genes that had not previously been known to have protooncogene function. The processing time from frozen tissue to cDNA sequence is five work days.

E. Possible modifications and variations on the best way

1. A partial digestion strategy is being implemented to recover more chimeric transcript sequence from loci which have CTTs that are too short to BLAST and/or contain low complexity or repetitive sequences. This will allow usable sequence to be recovered upstream of most CTTs currently listed as "unusable".

2. The HPT process has been implemented to recover host/virus junction fragments from integrants in the same transcriptional orientation as the target gene using minus strand primers from the 5' LTR. It is also possible to modify the procedure to recover host/virus junctions from integrants in the opposite orientation using plus strand primers from the 3' LTR.

3. The current procedure generates retroviral transcripts that run high in the gel so that novel host/virus junctions are clearly visible. It is also possible to remove, destroy, and/or inhibit the formation of retroviral transcripts.

4. The process claim can be generalized to include any method that uses a chimeric mRNA between a retrovirus and cellular gene to discover a gene of interest based on either an *in vivo* or cell culture bioassay.

F. Advantages and improvements over existing practice

The primary advantage and improvement over the existing state of the art is that the affected protooncogene is specifically recovered by the HPT process. Using DNA-based approaches, positional cloning is required to find the protooncogenes at loci that have not previously been characterized.

The following features are believed to be new:

1. First cDNA-based application of provirus tagging using PCR methods.

APPENDIX E

GLOSSARY

CTT	chimeric transcript tag
HPT	high-throughput protooncogene tagging
LTR	long terminal repeat
MMTV	mouse mammary tumor virus

APPENDIX F

CTTs FROM FIGURE 3

CTT	SAMPLE	BAND	SIZE	BLAST RESULTS
CTT0001	MM0001	A	101	*Wnt3*
CTT0002	MM0003	A	75	*Fgf3*
CTT0003	MM0006	A	46	*Fgf8*
CTT0004	MM0009	A	18	*Ins2/Igf2*
CTT0005	MM0010	A	79	novel
CTT0006	MM0029	A	51	*Fgf3*
CTT0007	MM0034	A	260	*Fgf3*
CTT0008	MM0038	A	45	*Myb*
CTT0009	MM0045	A	31	unusable
CTT0010	MM0064	A	46	novel
CTT0011	MM0065	A	201	novel
CTT0012	MM0078	A	2	unusable
CTT0013	MM0084	A	122	*Sp100*
CTT0014	MM0084	B	16	unusable
CTT0015	MM0092	A	22	*Wnt3a*
CTT0016	MM0094	A	322	*Fgf3*
CTT0017	MM0103	A	48	novel
CTT0018	MM0134	A	158	*Wnt1*
CTT0019	MM0250	A	2	unusable

CTT0001

CATGGCGAGA TTCTGTGTCC AAGCTGCCTC TACTCGTGAC ATTCCAAGAT GCCTCTGAGG
TGGGAACTGT GAAATAGGAC AGAGCCCCAC AGTCCCCTCT T

CTT0002

CATGGCAAGA TGGAGACTTT GTCTACCAGG GCCACTCCAA GCACCCAGCT GCATACAGGT
GGACTGGCTG TGGCC

CTT0003

CATGCTGGCT GTTCCTGCAG CCCAGCTACT GGGACAATCT GGAAAC

CTT0004

CATGTGCTCA ATCCATAG

CTT0005

CATGGGTCCC TGAAGGGTCT CTCCTTTAGC AAACCCCTGT ACAGTTGAAG TGATTTTTCA
GGTACCCATT GGTCTTAGC

CTT0006

CATGGCAAGA TGGAGACTTT GTCTACCAGG GCCACTCCAA GCACCCAGCT G

CTT0007

CATGCACACA AACTGGCCCT GAACTTTTGA CTTCCAGGCC TCTGCCTCTC TGCGCGCACA
CACACACTCG CACTCCTGTA TATGAAGCGT ATATGTGTTT CTCTGGGAAC TGTTTTTATC
AGGTGAAGTA CTTCCTTTGT TCTTGCTACC CACCTCCAGG GCTCCAGGAT CTCCAGACAG
CCAACCCTAA GACAGGCCCA GCTTCTCTGT ATCTCTGTGA TGAGAACCTT GGCATAGAGC
TGCCTCACCC TCGGGATAGG

CTT0008

CATGCCTCTG GAAAGTACCT TAAACATAGA ATCCCCTCCC TAGTG

CTT0009

CATGGTTTTT TTTTTTTTGA GTGTGTGTGT G

CTT0010

CATGCAGATT AAAGTACATA TATGTAAAAA ATAAAAATAA ATCTTT

CTT0011

CATGATAAGG TTAGAGTTTT GTGAGCCTCC TTAACCTTGC TCAGCAAGCG TTGGGCTCTT
GGCAGCCGAG CTGCCATCTT TCTCATCCCC GATAGAGCCA GCCGCCCTTG TCGTGTCTTG
AATAAGTTAG AGGAGGCATT ATAGAGCGGA CCTAAACATT TGCCTTGGAG CCTGAGGGAT

GGGGATTGGC TGAATGTGAA T

CTT0012

CA

CTT0013

CATGAATTCA TCACTGGTAA AATGTATGAA TTTCTTCTGA GACAGAGTCT TCTTATTGGC TTACACTTGC TTCGAGCGGA TGATTCTGCT GCTTCAGCCT CTTGAGATGC TCAGATATGT GC

CTT0014

CATGGATGCT ATTGGG

CTT0015

CATGAGAGGG TGCTTCAGGG TG

CTT0016

CATGCACACA AACTGGCCCT GAACTTTTGA CTTCCAGGCC TCTGCCTCTC TGCGCGCACA CACACACTCG CACTCCTGTA TATGAAGCGT ATATGTGTTT CTCTGGGAAC TGTTTTTATC AGGTGAAGTA CTTCCTTTGT TCTTGCTACC CACCTCCAGG GCTCCAGGAT CTCCAGACAG CCAACCCTAA GACAGGCCCA GCTTCCTCTG TATCTCTGTG ATGAGAACCT TGGCATAGAG CTGCCCTCAC CCTCGGGATA GGGCTTATGT TCCCCGGAAC GAGCCAGGCA CCTCAACAGC TCCTGGGGAG GAATAGGGGA CT

CTT0017

CATGAATTCC ACACCTCCAT CAAGGGTGTC TTCTCCAGTG AGCCCCGG

CTT0018

CATGCCTCCC TCAGCCTCCT CCCACCCCTT CCTGTCCTGC CTCCTCATCA CTGTGTAAAT AATTTGCACC GAAATGTGGC CGCAGAGCCA CGCGTTCGGT TATGTAAATA AAACTATTTA TTGTGCTGGG TTCCAGCCTG GGTTGCAGAG ACCACCCT

CTT0019

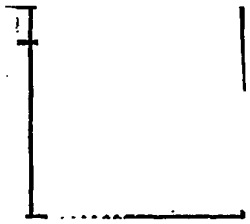
CA

APPENDIX G

NEW CANDIDATE PROTOONCOGENES

This group includes all novel CTT sequences ≥ 20 bp. Additional sequences are pending.

CTT LOCUS	SAMPLE	BAND	SIZE
CTT0005	MM0010	A	79
CTT0010	MM0064	A	46
CTT0011	MM0065	A	201
CTT0017	MM0103	A	48
CTT0020	MM0154	A	68



CTT0005

See Appendix F

CTT0010

See Appendix F

CTT0011

See Appendix F

CTT0017

See Appendix F

CTT0020

CATGCTAATG GAGTTTATTC TTAGGACTGC CTCCTGCATC CATTGATTGA CTTAAATATG TGCACACT

All cited references are expressly incorporated herein by reference.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 July 2002 (25.07.2002)

PCT

(10) International Publication Number
WO 02/057497 A3

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US02/01651

(22) International Filing Date: 18 January 2002 (18.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/262,724 18 January 2001 (18.01.2001) US

(71) Applicant and
(72) Inventor: MORRIS, David, W. [US/US]; 2841 Emerald Bay Drive, Davis, CA 95616 (US).

(74) Agents: SILVA, Robin, M. et al.; Flehr, Hohbach, Test, Albritton & Herbert, LLP, 4 Embarcadero Center, Suite 3400, San Francisco, CA 94111-4187 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
8 May 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/057497 A3

(54) Title: HIGH THROUGH-PUT CLONING OF PROTOONCOGENES

(57) Abstract: The present invention provides a process of identifying protooncogenes using high-throughput provirus tagging (HPT), e.g., by recovering host/virus junction sequences from chimeric transcripts containing both host and virus sequences.

INTERNATIONAL SEARCH REPORT

ıtional Application No
PCT/US 02/01651

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
BIOSIS, WPI Data, EPO-Internal, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	VAN LOHUIZEN M; VERBEEK S; SCHEIJEN B; WIENTJENS E; VAN DER GULDEN H; BERNS A.: "IDENTIFICATION OF COOPERATING ONCOGENES IN E-MU-MYC TRANSGENIC MICE BY provirus tagging" CELL, vol. 65, no. 5, 31 May 1991 (1991-05-31), page 737 XP001109573 NA US the whole document --- -/--	1

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 January 2003	06/02/2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Chambonnet, F

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	LI J ET AL: "Leukaemia disease genes: Large-scale cloning and pathway predictions" NATURE GENETICS, NATURE AMERICA, NEW YORK, US, vol. 23, no. 3, November 1999 (1999-11), pages 348-353, XP002223354 ISSN: 1061-4036 the whole document	1
X	BERNS, A. ET AL.: "Identification and characterization of collaborating oncogenes in compound mutant mice" CANCER RESEARCH, vol. 59, no. 7 suppl., 1 April 1999 (1999-04-01), pages 1773S-1777S, XP001109572 the whole document	1
X	JACOBS H.: "TCR-independent T cell development mediated by gain-of-oncogene fucntion or loss-of-tumor-suppressor gene function" SEMIN IMMUNOL. , vol. 12, no. 5, October 2000 (2000-10), pages 487-502, XP001148104 page 494, column 2	1
T	MIKKERS H. ET AL.: "High throughput retroviral tagging to identify components of specific signaling pathways in cancer" NAT GENET. 2002 SEP;32(1): , 19 August 2002 (2002-08-19) - September 2002 (2002-09), pages 153-159, XP001148101 the whole document	1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 02/01651

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

According to PCT Rule 13 ter.1.c., no specific sequence has been searched since the Sequence Listing as present in the description does not comply with WIPO Standard ST 25 prescribed in the administrative instructions under Rule 5.2. Since the Sequence Listing has not been furnished neither on paper form nor in machine readable form as provided for in the same instructions and since the applicant has not remedied the disclosed deficienceis within the time limit fixed in the invitation to PCT Rule 13ter.1.a.